

## Сравнение устойчивости программ филогенетической реконструкции к шуму во входных данных

Научный руководитель – Спириин Александр Сергеевич

*Логинковский Арсений Олегович*

*Студент (специалист)*

Московский государственный университет имени М.В.Ломоносова, Факультет  
биоинженерии и биоинформатики, Москва, Россия

*E-mail: arseniilog@list.ru*

Филогенетическая реконструкция — это определение эволюционных связей внутри набора последовательностей ДНК или белков. Результатом является представление эволюционных связей в виде филогенетического дерева. Достаточно часто возникает необходимость использования фрагментированных или сильно зашумленных данных для филогенетической реконструкции. Выяснение зависимости частоты ошибочного построения филогенетического дерева от степени зашумлённости данных может представлять практический интерес, так как позволит избежать наиболее распространенных ошибок, в частности, притяжения длинных ветвей [2].

В работе А.Галицыной [1] было проведено исследование устойчивости к шуму различных алгоритмов реконструкции филогении нуклеотидных последовательностей (рРНК). Задачей данной работы было провести аналогичное исследование для филогении белков.

Проводилась реконструкция филогенетического дерева 45 протеобактерий по объединённому выравниванию 29 рибосомальных белков. Для каждого из 5 классов протеобактерий (Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, Epsilonproteobacteria) было выбрано по 9 организмов.

Проверялась правильность выделения тремя алгоритмами (Фитча - Марголиаша, объединения соседей и сбалансированной минимальной эволюции) каждого из пяти классов, а также двух филогенетических групп (beta+gamma и alpha+beta+gamma).

С помощью пакета BioPython из общего выравнивания длиной 5647 генерировались выравнивания из небольшого (несколько десятков) количества случайно выбранных позиций. Программой protdist из пакета PHYLIP [3] для каждого сгенерированного выравнивания строилась матрица расстояний, с которой в дальнейшем работали сами алгоритмы филогенетической реконструкции. Выбор небольшого количества позиций равносильно внесению случайного шума в матрицу расстояний.

Для каждой из семи групп организмов и каждого алгоритма проверялась зависимость частоты ошибок при реконструкции группы от длины сгенерированного выравнивания.

Показано, что при восстановлении различных филогенетических групп наиболее устойчивыми к шуму оказываются разные алгоритмы. Так, класс Betaproteobacteria заметно устойчивее выделяется алгоритмом Фитча - Марголиаша, в то время как для класса Deltaproteobacteria этот алгоритм, наоборот, показывает наихудшие результаты из рассмотренных. Планируется провести аналогичный анализ для различных подмножеств 45 организмов с целью обнаружить особенности (такие, как длинные ветви), влияющие на относительную устойчивость различных алгоритмов к шуму.

### Источники и литература

- 1) А. Галицына. Сравнение устойчивости программ реконструкции филогении к шуму во входных данных // Материалы XX Международного молодежного научного форума «Ломоносов-2013»: [https://lomonosov-msu.ru/archive/Lomonosov\\_2013/2064/60120\\_3605.pdf](https://lomonosov-msu.ru/archive/Lomonosov_2013/2064/60120_3605.pdf)

- 2) J.Bergsten. A review of long-branch attraction // Cladistics, 2005 21(2): 163-193
- 3) PHYLIP: <http://evolution.genetics.washington.edu/phylip.html>